周期，是一套完整的新闻事件信息数据解决方案。这点可以保证新闻事件数据在多个技术系统间实现自由流转，而并不需要转换，大大降低了数据使用和管理的成本，也降低了相关系统的开发难度。

其次，REML 标准与现有中文新闻信息国家标准（例如中文新闻信息置标语言，CNML）之间有良好的兼容性，对已经采用了 CNML 标准的单位、系统来说都不存在任何颠覆性技术风险。

再次，REML 标准采用了面向知识的语义技术，不仅可以描述新闻事件数据本身，还可以与外部知识库进行关联，进而实现新闻事件数据的语义化分析处理，甚至以此为基础建立新闻事件知识库。

2.3 应用及前景

REML 标准的研发过程采用了更先进的“研制－应用”双线螺旋式推进模式。该模式的优势显而易见，标准在研制过程中就充分考虑了实际应用的需要，并且通过实际应用对标准进行充分的验证和修正，避免了标准高高在上，与实际应用脱节的问题。

随着融媒体服务、知识服务和智能服务的概念在新闻领域逐渐兴起和发展，新闻生产和利用逐渐向数据融合、知识融合的方向快速前进。新闻事件数据作为行业中最重要的知识和资产之一，必将成为各新闻媒体单位关注的重点。而 REML 标准的颁布实施，则从标准层面保证了新闻事件从生产端开始就是以高质量结构化的形式保存，并且能够有效交换和传输，最后还能以语义数据的形式成为知识资产，为构建基于新闻事件知识的智能服务提供基础。

《中文新闻图片内容描述元数据规范》国家标准研制

（张弛，副教授，硕士生导师，中国传媒大学计算机学院计算机系副主任，主要研究方向为媒体信息数据化技术，从事自然语言处理、数据挖掘、知识库等方面的研究）

所谓“一图胜千言”，视觉信息在人类社会信息生产和消费中始终扮演着重要角色。在融合媒体不断发展的背景下，图片凭借其信息量大、视觉吸引力强、数据量小等优势，在两微一端应用、精品内容生产等方面不可或缺。资源价值的发挥有赖于资源的有效流通，而流通的前提是资源发现。图片元数据是图片特征描述的核心，也是图片发现、流通的基础。

国家标准——《中文新闻图片内容描述元数据规范》（GB/T 35311–2017）于 2017 年 12 月底正式发布，填补了我国新闻图片应用标准的一项空白，也是世界范围内第一个以图片视觉内容和语义特征描述为主的图片元数据规范，专业性强，结构简明，有较好的可扩展性和兼容性，具有较高的应用价值。

一、标准研制背景

视觉内容已经成为连接媒体、品牌以及消费者的最高效的传播媒介[1]。图片是视觉内容存储和传播的重要载体。随着互联网技术和数码设备的不断发展和普及，图片数量剧增、质量不断提升、需求也日益扩大。多家国内外大型商业图片公司、媒体机构等拥有千万级甚至亿级规模的高质量图库，为新闻、出版、印刷、广告、营销、素材制作等提供图片素材。各类特色图片库以及自用为主的图片系统更是不胜枚举。

图片特征描述在图片存储、管理、检索、交易交换、智能推荐等生命周期各个阶段都具有重要作用。没有规矩不成方圆，我们需要图片元数据标准来规范图片生产的各个环节。原有图片元数据相关标准各有侧重，且大都缺乏对图片视觉内容和语义特征的系统化规约，未能很好地满足日益增长和不断变化的图片标注需求。目前大部分图片应用系统从自身业务需求出发，自定义了若干图片元数据元素来描述图片内容特征，这些私有化的元数据元素难以在不同业务间流转交换，制约了图片价值的发挥。

在此背景下，新华通讯社提出了《中文新闻图片内容描述元数据规范》国家标准编制申请，经全国中文新闻信息标准化技术委员会（以下简称新标委）上报国家标准委，被国家标准委批准列入“2014 年第一批国家标准计划项目”。该标准重点关注图片视觉内容和语义特征描述，可用于图片数据的采集、编辑、存储、发布、检索、交换等处理环节。

二、标准编制过程

2015 年 5 月新标委秘书处牵头组建标准编制组，确定了以媒体机构及大型商业图片库为主要调研对象、参考国外相关机构案例和国内外相关标准规范的工作路线。编制组历时一年，先后对多家报社、商业图片公司等单位进行实地调研，并充分借鉴了国内外图片描述相关标准，形成了标准的征求意见稿。在经过多次的意见反馈和修改后，于 2016 年底提交送审稿到新标委秘书处。经

新标委委员函审投票后，在 2017 年 4 月召开送审稿审查会。再次修改后形成报批稿，上报国家标准委批准发布。国家标准《中文新闻图片内容描述元数据规范》于 2017 年 12 月底由国家标准委正式对外公布，可在国家标准全文公开系统 [2] 中查询。

该标准的核心目标是更好地服务于图片生产和消费的各个环节，突出标准的实用性，因此，编制组在标准研制初期作了大量的调研工作。

调研对象包括两类，一是现有相关国内外标准，二是标准主要应用对象。标准编制组首先对国内外标准，包括 IPTC 图片元数据规范 [3]、EXIF 规范 [4]、CNML 标准 [5] 等进行研究。研究结果发现，每个标准侧重点不同。IPTC 图片元数据规范主要定义图片描述型元数据项，几经改版后，补充了若干图片内容描述的元数据项，但较为零散，在图片内容语义特征描述方面不够全面。EXIF 规范主要涉及图片技术型元数据。CNML 标准主要对象是新闻稿件，图片内容相关元数据定义相对较少。图片元数据主要应用对象是图片库系统。在对图片库网站、媒体机构图片库系统的调研中发现，个性化、私有化图片元数据应用现象普遍存在，这其中部分元数据仅仅名称不同，含义却相近，部分元数据具有普遍应用价值。此外，旧的图片库系统越来越难以满足应用需求，新的图片库系统设计时在图片编目方面也存在较多困惑。标准调研过程及意见征集过程中得到了图片各领域专家的积极配合和大力支持，使标准内容不断得到丰富和完善。

三、标准主要内容与特色

（一）标准内容概览

本标准的主要内容是对图片元数据元素做出准确定义，尤其突出对图片视觉内容特征和语义信息的描述。标准以编辑类图片和创意类图片为主要对象，定义了 40 个元数据元素，分别归属到公共元数据、编辑类图片元数据和创意类图片元数据三个部分。

本标准主要从四个角度进行元数据元素的筛选和梳理。一是“形”，即视觉表现，主要指画面中的主要事物及环境信息等；二是“意”，即图片主题，主要指概念、专题、分类等；三是“技”，即创作手法，主要指画面构图、摄影技术信息。四是图片基本逻辑属性，主要指图片标识、新闻图片标题和图说等信息。标准具体内容详见国家标准全文公开系统 [2]。

（二）标准特色

1. 从应用中来

实际应用中，图片发现可分为检索和推荐两类。图片检索主要包括关键词检索、类别检索、专题检索和属性过滤筛选等。关键词检索对于创意类图片侧重主题和

画面内容，对于编辑类图片侧重新闻人物、地点和事件信息。用于检索结果集合筛选过滤的属性主要集中在图片色彩、片幅、尺寸、人物特征等。图片推荐包括热门、相似、个性化以及编辑推荐等，其中相似推荐大多以图片色彩和主题相似分析为基础。标准中定义的图片内容元数据元素大部分来源于此。

2. 有详有略

标准中对人物特征描述和摄影技术进行了较为详细的定义。人数、性别、年龄、民族、行为、表情、服饰、人物关系等人物信息，图片色彩、画幅方向、景别、色调、构图方式、摄影技巧等摄影技术信息，对搜索结果过滤精化、图片自动分析与推荐都具有重要作用。

标准中对图片专题、自定义分类、画面内容等未做细化规定，以适应多种不同业务场景的需要，可由用户自定义扩展，提高标准的适用性。

3. 强调规范化标注

标准中定义了 22 个可扩展的受控词表，用于规范 20 个元数据元素的标注，减少主观标注的不一致性，也便于支持图片检索时的结果集过滤。

4. 层次化、体系化

标准编制组首先建立了一套图片元数据体系架构，根据语义抽象程度不同，依次分为物理层、逻辑层、内容层和主题层。原有图片元数据描述规范定义的元数据项集中在物理层和逻辑层。本标准主要定义了内容层和主题层元数据，这相比传统图片基本属性编目具有更高的辨识度和利用价值。

标准编制过程中，得到多位图片领域专家的指导与支持，编制组从图片特性、新闻图片编辑、图片应用现状和发展等多个角度，不断修正完善，形成了层次简明清晰、专业性较强、体系较为完整的图片内容描述元数据规范。

四、标准应用思考

标准来自于应用，服务于应用，更需要在应用中不断完善。标准应用中有两个方面的问题需要认真思考和应对。

第一是存量图片标注和已有图片库系统改造问题。随着图片业务的不断变化和升级，目前很大一部分图片库系统面临升级改造问题。同时，存量图片数量巨大，且已经基于原有标准和自定义元数据进行了标注。升级图片库系统的同时，如何花费尽可能少的工作量完成存量图片元数据信息向新的图片元数据规范的转换，是该标准推广应用中的重要问题。图片库升级以及存量图片元数据迁移是必要的，符合软件生命周期规律，也是技术和业务发展的要求。较可行的方案是图片系统设计人员整理出存量图片已标注的元数据与新标准中元数据的对应关系，区分出直接对应、间接对应、无法对应等不同情况，最大程度的采用自动化方式实现部分元数据的迁移转换，保留已标注元数据内容，再逐步完成新的元数据的标注工作。

第二是图像智能处理技术与基于文本的图片元数据规范在应用中的关系问题。随着机器学习和人工智能技术的快速发展，计算机在图像处理、图像识别、图像理解等方面的智能化程度和水平不断提高，已由特定领域应用迈入通用领域应用，但同时我们不难发现，基于文本的图片特征描述仍然在图片存储、编辑、管理、检索和交换中扮演主要角色。二者之间不是对立关系，而是相辅相成、互为补充的关系。图片元数据标准在应用中可充分利用计算机自动处理技术，提高标注效率和质量；基于文本的图片元数据标注结果具有丰富的图片规格、内容、主题等特征，可在图片自动处理与智能推荐等环节发挥重要作用。双管齐下、各尽其能，使图片更好地为社会生产生活服务。

五、总结与展望

国家标准《中文新闻图片内容描述元数据规范》从新闻编辑用图和创意用图两个应用角度，规定了中文新闻图片视觉内容和语义信息的元数据元素集合，有利于跨领域、跨应用的图片交换，更好地服务于图片采编发和分析推荐等业务环节。标准中未涉及图片版权元数据规约，应用中可遵循其他版权元数据标准。刚刚发布的这一版标准中尚未对数字图表、漫画、动图等其他类型图片的个性化元数据做出规定，但标准中已定义的元数据元素可适用于这些类图片的标注。

该标准研制过程中得到了很多学者、领域专家的帮助和支持，编制组力求在完整性、专业性、实用性等方面达到相对平衡，但不断完善才能保持其生命力。希望标准能够得到越来越多的应用，并在应用中发现不足，促进标准的升级改版，从而更好地为应用服务。

参考文献

[1] 任悦，曾璜编著 .《图片编辑手册》第四版 . 中国摄影出版社 . 2015 年 9 月

[2] 国家标准全文公开系统 . http：//www.gb688.cn/bzgk/gb/index

[3]IPTC Photo Metadata Standard. http：//www.iptc.org/std/photometadata/specification/IPTC-PhotoMetadata

[4]EXIF Image Metadata. https：//www.exif.org/image-metadata.html

[5] 全国中文新闻信息标准化技术委员会 . 国家标准 GB/T 20092-2013《中文新闻信息置标语言》